Chapter 29

DNA Barcoding Techniques for Protists

Amit Gamit and Dhruti Amin

Abstract

The study of free-living protist populations in soil is gaining momentum due to their crucial role in soil ecosystems, particularly in carbon and nitrogen cycles. Nevertheless, there are notable gaps in our understanding of the abundance and functions of protists in soil. When it comes to assessing soil microorganisms, DNA extraction from environmental samples is a standard practice. However, soil eukaryotic microbes, particularly protists, have not received as much attention as other microorganisms. A similar-looking but genetically different organism cannot be distinguished by phenotypic features alone. Molecular biology tools such as DNA barcoding enable the identification of different organisms based on their genetics. This protocol chapter deals with general procedures for the sample collection, and morpho-molecular identification of protist which helps in building the DNA barcoding library of known species of protists. Eventually, this makes it easy for knowing the diversity and to study the applications of protists widespread.

Key words Species identification, Molecular tool, DNA barcoding, Diversity

1 Introduction

A protist is a heterogeneous group of microbial eukaryotes consisting of protozoa, algae, and slime molds. Protozoa are believed to be the first eukaryotes, originating from a primary endosymbiotic event in which archaeal prokaryotes engulfed *Proteobacteria* which eventually became mitochondria, revolutionizing physiology and altering life processes and the diversity of organisms [1–3]. They emerged after archaea and bacteria but before plants, animals, or fungi. Due to their single-celled nature, identifying and describing new protist species has been challenging. As a result, many protist groups have only a small number of formally classified species. Despite this, certain protist groups play crucial roles as pathogens, environmental indicators, and indicators of past environmental conditions. Using molecular techniques like DNA barcoding could help to classify protists, compensating for their lack of distinct physical features. However, the immense diversity within this group has made this challenging. The genetic differences observed

N. Amaresan and Komal A. Chandarana (eds.), *Practical Handbook on Soil Protists*, Springer Protocols Handbooks,
https://doi.org/10.1007/978-1-0716-3750-0_29,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2024

between and within major protist groups are much greater than those found in the three main multicellular kingdoms. While no single set of molecular markers works for all lineages, an international working group is nearing a solution. Once established, a universal DNA barcode for protists, along with group-specific barcodes, will revolutionize taxonomic research and lead to various practical applications [4]. It is becoming increasingly important to study protist populations in the soil as they play an integral role in soil ecosystems [5, 6].

Protists play a pivotal role in shaping microbial communities and influencing the dynamics of soil ecosystems through predation [7]. Approximately 85% of the entire soil protist community are predators, targeting a diverse range of microbes including bacteria, fungi, nematodes, and even fellow protists [7, 8]. This predatory activity holds vital significance for plants, as it leads to the release of nutrients trapped within the cells of rhizosphere microbes, ultimately benefiting plant nutrition [7]. Notably, protist predation exhibits a high degree of selectivity, allowing them to target specific prey, including pathogenic microbes, thereby contributing to enhanced plant protection [9]. Harnessing this mechanism offers the potential for implementing sustainable plant protection strategies by integrating soil protists into belowground ecosystems [7].

A DNA barcode identifies a species by using a short segment of DNA from a specific gene or genes. By using DNA barcodes, nonexperts can objectively identify species from small, damaged, or industrially processed materials. A "DNA barcode" is a unique DNA sequence pattern that can potentially identify every living thing in the world, much like the unique bars in a universal product code (UPC). According to DNA barcoding, an individual sequence can be compared with a reference library of DNA sequences to uniquely identify an organism to species, just as a supermarket scanner compares items in its stock with the reference database of the UPC barcode. By using computer programs, short DNA barcodes, about 700 nucleotides in length, can be unambiguously analyzed from thousands of specimens quickly [10]. There is a common assumption that genetic variation between species is greater than within species. This approach is particularly useful when phenotypic plasticity is a concern, morphology keys are not available or unknown, or cryptic species are present. Another primary purpose of this approach is to discover new species [11].

Different gene regions are used to identify organismal groups by barcoding. The most commonly used barcode region for animals and some protists is a portion of the cytochrome c oxidase I (COI or COX1) gene found in mitochondrial DNA. Other genes suitable for DNA barcoding are the internal transcribed spacer (ITS) rRNA, often found in fungi, and RuBisCO, often found in plants [12–14]. Microorganisms are detected by different gene regions. The 16S rRNA gene, for example, is widely used for

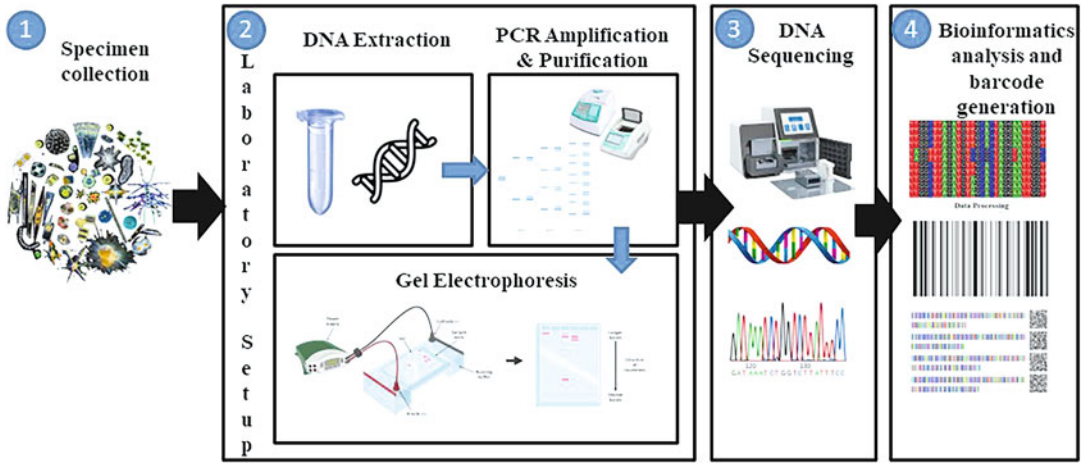

Fig. 1 Steps involved in DNA barcoding of protist (1) specimen collection; (2) DNA Extraction-PCR amplification and purification; (3) DNA sequencing; (4) bioinformatics analysis and barcode generation

identifying prokaryotes. In contrast, the 18S rRNA gene is mainly used for detecting microbial eukaryotes. These gene regions are chosen because they have less intraspecific (within species) variation than interspecific (between species) variation, which is known as the "barcoding gap" [15].

Sanger sequencers produce a chromatogram (also known as an electropherogram or trace file), from which the DNA sequence is derived. Common file formats for chromatograms are ABI (created by the Applied Biosystems sequencers) and SCF (created by platforms such as Beckman or Li-Cor). These files require specific software to be viewed and analyzed. Free software options include BioEdit49. The DNA database is then searched using the sequencing results. Close matches help identify species that are already in the database. Certain barcodes, however, may be entirely new, and identification may depend on placing the unknown species in a phylogenetic tree with near relatives. Novel DNA barcodes can be submitted to GenBank. An overview of the DNA barcode generation process for protists is shown in Figs. 1 and 2.

2 Materials

2.1 DNA Extraction

1. Established culture
2. DNA extraction kit
3. Microcentrifuge tubes (1.5 mL), micropipettes and tips (2–1000 μL), microcentrifuge tube rack
4. Microcentrifuge, vortex, water bath, or heating block at 65 °C and 57 °C

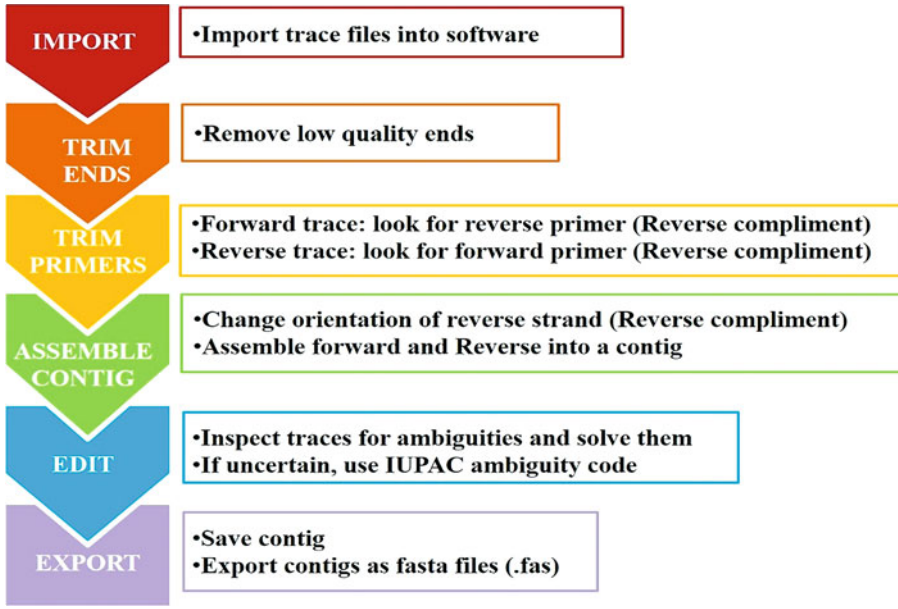
IMPORT
•Import trace files into software
TRIM ENDS
•Remove low quality ends
TRIM PRIMERS
•Forward trace: look for reverse primer (Reverse compliment)
•Reverse trace: look for forward primer (Reverse compliment)
ASSEMBLE CONTIG
•Change orientation of reverse strand (Reverse compliment)
•Assemble forward and Reverse into a contig
EDIT
•Inspect traces for ambiguities and solve them
•If uncertain, use IUPAC ambiguity code
EXPORT
•Save contig
•Export contigs as fasta files (.fas)

Fig. 2 DNA sequence editing for the generation of barcodes

2.2 PCR Amplification

1. Specific primers
2. Template DNA
3. 1× PCR buffer
4. 2 mM $MgCl_2$
5. 1 U AmpliTaq DNA polymerase
6. 100 μM deoxynucleoside triphosphates
7. 0.4% bovine serum albumin
8. Micropipettes and tips (2–1000 μL), PCR tubes, PCR rack, ice block
9. Thermal cycler

2.3 Gel Electrophoresis and Purification

1. 1% agarose in 1X TAE
2. 100-bp ladder
3. PCR products
4. 1X TAE buffer (~300 mL per gel)
5. EtBr (ethidium bromide)
6. Bromophenol blue (or gel-loading dye)
7. Latex gloves, masking tape
8. Microwave, gel-casting tray and comb, gel electrophoresis chamber and power supply, digital camera or photodocumentary system, UV or LED transilluminator and eye protection

2.4 DNA Sequencing

1. Sanger DNA sequencer/high-throughput sequencing (HTS) platforms (Sequel, Sequel II)

2.5 Bioinformatics Analysis and Barcode Generation

1. Computer
2. High-speed Internet

3 Methods

3.1 DNA Extraction

1. Collect the enriched growth of protist in a microcentrifuge tube, and carry out DNA extraction either using DNA extract kit according to manufacturer's instruction [16], or follow the manual method described in this book (Chapters 26 and 27).
2. Check out the efficacy of the DNA by electrophoresis on 1% agarose gel stained with ethidium bromide.

3.2 PCR Amplification

Use primers which are specifically designed for 18s rRNA gene/or any marker gene of protist phylum. Carry out PCR amplification using any of the below mention primer set/pairs (Table 1) (s*ee* **Note 1**).

1. Perform PCR amplification using 50 μL reaction mixtures which include 1X PCR buffer, 2 mM $MgCl_2$, and 1 U Ampli-Taq DNA polymerase, 100 μM deoxynucleoside triphosphates, 0.2 μM forward and reverse primers, and 0.4% bovine serum albumin.
2. Add 2 μL of template DNA extracted from the protist samples.
3. PCR settings for mitochondrial genome sequence (mtDNA): initial denaturation at 94 °C for 60 s followed by denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, and extension at 72C for 30 s with final extension at 72 °C for 4 min for 35 amplification cycles [20].
4. PCR settings for the ITS fragments: initial denaturation at 95 °C for 5 min, followed by denaturation at 95 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 90 s with a final extension at 72 °C of 5 min for 35 amplification cycles [22].
5. PCR settings for COX1: hold at 94 °C for 4 min; denaturation at 94 °C for 30 s, annealing at 55 °C for 1 min, and extension at 72 °C for 105 s with a final extension at 72 °C for 10 min for 35 amplification cycles [23].
6. PCR settings for SSU rDNA: hold at 94 °C for 4 min, denaturation at 94 °C for 1 min, annealing at 55 °C for 90 s, and extension at 72 °C for 3 min; and hold at 72 °C for 10 min for 35 amplification cycles [23].

Table 1
List of primer schemes for barcode markers commonly used for protists

Target	Primer	References
3′ fragment of the SSU rRNA (*see* **Note 2**) ITS region (*see* **Note 2**)	s12.2 (5′-GAT YAG ATA CCG TCG TAG TC-3′) sB (5′-TGA TCC TTC TGC AGG TTC ACC TAC-3′) Gfor (5′-GGGATCCGTTTCCGT AGGTGAACCTGC-3′) Grev (5′-GGGATCCATATGCTTAAGTTCAGCGGGT-3′)	[17, 18]
A 5′ fragment of COI gene ca 650 bp in length (*see* **Note 2**)	LCO1490F (5′-GGTCAACAAATCATAAAGATATGG-3′) HCO2198R (5′TAAACTTCAGGG TGACCAAAA-3′)	[18, 19]
Mitochondrial genome sequence (mtDNA)	EGcox1-F2 (5′-GCTCATTCAGGAGGTTCTGT-3′) EGcox1-R4 (5′-GATAAAACAATACCTGTTAA-3′) EGcox1-F3 (5′-ACTATGATTATTGCTGTTC-3′) EGcox1-R5 (5′-ACTAAGTAATCAGTTTCTGC-3′) EGcox1-16274F (5′-TGCAATTGCTTCATTTGGTAC-3′) EGatp4-16959R (5′-TGCCGATTTCGCATCAATAAG-3′)	[20, 21]
3′ of 18S rRNA gene 5′ of 28S rRNA gene	ITSF (5′-CGTAACAAGGTTTCCGTAGG-3′) ITSR (5′-TCCTCCGCTTACTGATATGC-3′)	[22]
11080 of *P. aurelia* mt genome 10139 of *P. aurelia* mt genome	CauCoxF (5′-TCAGGAGCTGCMTTAGCTCCYATG-3′) CauCoxR (5′-TARTATAGGATCMCCWCCATAAGC-3′)	[22]
11058 of *P. aurelia* mt genome 10176 of *P. aurelia* mt genome	CoxL11058 (5′-TGATTAGACTAGAGATGGC-3′) CoxH10176 (5′-GAAGTTTGTCAGTGTCTATCC-3′)	[22]
COX1 gene	F (5′-ATGTGAGTTGATTTTATAGAGCAGA-3′) R (5′-GGDATACCRTTCATTTT-3′)	[23]
SSU rDNA	FA (5′-CAACCTGGTTGATCCTGCCAGT −3′) RC (5′-TTGGTCCGTGTTTCAAGACG-3′)	[23–25]

3.3 Analyze PCR Products by Gel Electrophoresis

1. Take 5 μL of PCR products for gel electrophoresis.
2. Use 1.5% agarose dissolved in TAE buffer and 1 Kb Plus DNA ladder (Invitrogen) as a marker to verify the presence of PCR products.
3. Run the gel for approximately 30 min at 80–100 V.
4. View the gel using UV or LED transillumination. Photograph the gel using a digital camera or photodocumentary system.
5. Only PCR products with single and clear bands to be used for Sanger sequencing/high-throughput sequencing (HTS) platforms (Sequel, Sequel II) to ensure amplification success.

3.4 DNA Sequencing [26]

1. Once PCR amplicons are generated, the next step involves their sequence characterization.
2. Properly label the sample and send it for DNA sequencing. High-throughput sequencing (HTS) platforms (Sequel, Sequel II) are used for DNA sequencing.
3. The sequencing facility will return results as sets of files generated by the DNA sequencer (trace files/chromatograms with Sanger sequencing), each file corresponding to a sample.

3.5 Bioinformatics Analysis [27]

1. For reference library building, upload these sequence files to BOLD (Barcode of Life Database) to preserve the raw data.
2. Upload the obtained sequence file on the BOLD [28]; it will allow matching with similar sequence which is already reported and helps to identify the query sequence (*see* **Note 3).**
3. Steps for data upload are detailed in the BOLD Handbook (*see* **Note 4).**
4. Use a field guide or taxonomic key to identify the specimen as precisely as possible: kingdom > phylum > class > order > family > genus > species.
5. Check to see if specimen is represented in the Barcode of Life Database, BOLD (www.boldsystems.org) or GenBank (www.ncbi.nlm.nih.gov).
6. Search by entering genus and species names in the search bar at top right. If the species is represented in the database, the "Taxonomy Browser" will list the number and sources of specimen records.
7. Click on "Download Public Sequences" for a FASTA file of available barcode sequences.
8. Click "Taxonomy Browser" at the top left to explore barcode records by group.

4 Notes

1. The forward and reverse primers can be designed from reference sequence available on GenBank.
2. The PCR conditions were as described by Fahrni et al. [29].
3. The Barcode of Life Data System (BOLD) is an informatics workbench aiding the acquisition, storage, analysis, and publication of DNA barcode records. By assembling molecular, morphological, and distributional data, it bridges a traditional bioinformatics chasm.
4. Users can verify the accuracy of sequence editing by downloading and analyzing these files. HTS platforms do not generate chromatograms which is why only sequence files in text format (FASTA) are uploaded.

References

1. Margulis L (1993) Symbiosis in cell evolution, 2nd edn. Freeman, NewYork
2. Fenchel T (2002) Origin and early evolution of life. Oxford University Press, Oxford
3. Foster K (2003) Making a robust bimolecular time scale for phylogenetic studies. Protist 154: 43–55
4. Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C et al (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol 10:e1001419
5. Countway P, Gast R, Savai P, Caron D (2005) Protistan diversity estimates based on 18s rdna from seawater incubations in the western North Atlantic. J Eukaryot Microbiol 52:95–106
6. Martínez J (2013) Bacterial pathogens: from natural ecosystems to human hosts. Environ Microbiol 15(2):325–333
7. Gao Z, Karlsson I, Geisen S, Kowalchuk G, Jousset A (2019) Protists: puppet masters of the rhizosphere microbiome. Trends Plant Sci 24:165–176
8. Oliverio AM, Geisen S, Delgado-Baquerizo M, Maestre FT, Turner BL, Fierer N (2020) The global-scale distributions of soil protists and their contributions to belowground systems. Sci Adv 6:eaax8787
9. Amacker N, Gao Z, Hu J, Jousset AL, Kowalchuk GA, Geisen S (2022) Protist feeding patterns and growth rate are related to their predatory impacts on soil bacterial communities. FEMS Microbiol Ecol 98:fiac057
10. Mora C, Tittensor D, Adl S, Simpson A, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biol 9:e1001127
11. Hebert PDN (2004) Identification of birds through DNA barcodes. PLoS Biol 2:312–319
12. Dawson S, Pace N (2002) Novel kingdom level eukaryotic diversity in anoxic environments. Proc Natl Acad Sci U S A 99:8324–8329
13. López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409:603–607
14. Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic "kingdoms"? Pitfalls and limitations of environmental DNA surveys. BMC Biol 2:13
15. Kim E, Harrison J, Sudek S, Jones M, Wilcox H (2011) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. Proc Natl Acad Sci U S A 108:1496–1500
16. Plassart P, Terrat S, Thomson B, Griffiths R, Dequiedt S, Lelievre M et al (2012) Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. PLoS One 7: e44279
17. Dyková I, Boháčová L, Fiala I, Macháčková B, Pecková H, Dvořáková H (2005) Amoebae of the genera Vannella Bovee, 1965 and Platyamoeba page, 1969 isolated from fish and their phylogeny inferred from SSU rRNA gene and ITS sequences. Eur J Protistol 41:219–230
18. Kosakyan A, Heger TJ, Leander BS, Todorov M, Mitchell EA et al (2012) COI barcoding of nebelid testate amoebae (Amoebozoa: *Arcellinida*): extensive cryptic diversity and redefinition of the *hyalospheniidae schultze*. Protist 163(415):434
19. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299
20. Hagino K, Bendif EM, Young JR, Kogame K, Probert I, Takano Y et al (2011) New evidence for morphological and genetic variation in the cosmopolitan coccolithophore emiliania huxleyi (prymnesiophyceae) from the COX1b-ATP4 genes 1. J Phycol 47:1164–1176
21. Virginia M, Puerta S, Bachvaroff T, Delwiche C (2004) The complete mitochondrial genome sequence of the haptophyte *Emiliania huxleyi* and its relation to heterokonts. DNA Res 11:1–10
22. Barth D, Krenek S, Fokin SI, Berendonk TU (2006) Intraspecific genetic variation in Parame cium revealed by mitochondrial cytochrome C oxidase I sequences. J Eukaryot Microbiol 53:20–25
23. Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N et al (2007) Barcoding ciliates: a comprehensive study of 75 isolates of the genus Tetrahymena. Int J Syst Evol Microbiol 57:2412–2425
24. Jerome CA, Lynn DH (1996) Identifying and distinguishing sibling species in the Tetrahymena pyriformis complex (*Ciliophora, Oligohymenophorea*) using PCR/RFLP analysis of nuclear ribosomal DNA. J Eukaryot Microbiol 43:492–497
25. Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71:491–499

26. Hebert P, Braukmann T, Prosser S (2018) A sequel to Sanger: amplicon sequencing that scales. BMC Genomics 19:219
27. Saunders GW, McDevit DC (2012) Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. In: Kress W, Erickson D (eds) DNA barcodes, Methods in molecular biology, vol 858. Humana Press, Totowa, pp 207–222
28. www.boldsystems.org
29. Fahrni JH, Bolivar I, Berney C, Nassonova E, Smirnov A, Pawlowski J (2003) Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. Mol Biol Evol 20:1881–1886